

ABSTRACT OF THE DISCLOSURE

A known method for selecting words (or word sequences), which is an important aspect of information retrieval, involves the problems of inability to eliminate high-frequency common words and of often arbitrary setting of the threshold value for dividing important and unimportant words. These problems are solved by normalizing the difference between the word distribution in a subset of all documents containing a word to be extracted (or a subset of said document set) and the word distribution in the set of all documents with the number of words in the said subset of all documents containing the word as a parameter, and the accuracy of support to information retrieval is thereby enhanced.